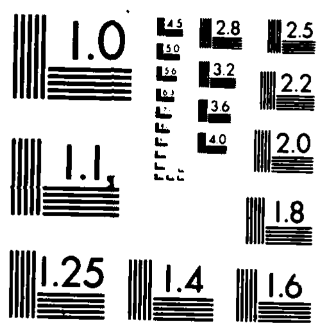
AD-A129 930 VOICE RECOGNITION ACCURACY: WHAT IS ACCEPTABLE?(U) 1/1
NAVAL POSTGRADUATE SCHOOL MONTEREY CA G K POOCK ET AL.
NOV 82 NPS55-82-030 MIPR-TB-024

UNCLASSIFIED F/G 17/2 NL

END
DATE
FILMED
7-83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

NPS55-82-030

NAVAL POSTGRADUATE SCHOOL

Monterey, California

VOICE RECOGNITION ACCURACY:

WHAT IS ACCEPTABLE?

by

G. K. Poock
E. F. Roland

November 1982

Approved for public release; distribution unlimited.

Prepared for:
Naval Electronic Systems Command
Washington, D. C. 20360
and
9th Infantry Division
Fort Lewis, WA 98433

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral J. J. Ekelund
Superintendent

D. A. Schrady
Provost

This work was performed by the authors at the Naval Postgraduate School. Professor Poock has been investigating the potential for VOICE recognition/input into both Navy and Army systems (Navy Document Number N00039WRDX017 and Army MIPR TB-024). E. F. Roland has also performed work as a contractor to NPS for Professor Poock under "Research and Development Study of the Feasibility of Using Computer Voice Entry" under NPS Contract N00228-82-C-6418.

Individual reports have been prepared for each sponsor on studies pertinent to their work. The enclosed work was not required by either sponsor nor funded by either sponsor specifically. We did this work in our spare time as we felt it was very important. Because we feel the topic is very generic to both Army and Navy, we have prepared the report for both.

Gary K. Poock

G. K. Poock, Professor
Department of Operations Research

Ellen F. Roland

E. F. Roland
Rolands and Associates

Reviewed by:

K. T. Marshall

K. T. Marshall, Chairman
Department of Operations Research

Released by:

William M. Tolles

William M. Tolles
Dean of Research

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPS55-82-030	2. GOVT ACCESSION NO.: AD-A129930
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): VOICE RECOGNITION ACCURACY: WHAT IS ACCEPTABLE?	5. TYPE OF REPORT & PERIOD COVERED: Technical
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): G. K. Poock E. F. Roland	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Postgraduate School Monterey, CA 93940	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: N0003982WRDX017 and Army MIPR TB-024
11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Postgraduate School Monterey, CA 93940	12. REPORT DATE: November 1982
	13. NUMBER OF PAGES: 51
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 1. Naval Electronic Systems Command Washington, D.C. 20360 2. 9th Infantry Division Fort Lewis, WA 98433	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

VTAG
Voice Recognition
Automatic Word Recognition
User Acceptability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A research experiment was conducted to investigate how accurate a voice recognition system must be for daily production use. Specifically, the purpose of the research was to establish the percentage accuracy level at which a user becomes frustrated and decides not to use a voice recognition device. The experiment consisted of controlling the perceived recognition accuracy of a voice recognition system and then collecting data through the use of a

questionnaire from the experimental users on the acceptability of the equipment. The experiment was not totally successful for a variety of reasons. This paper will discuss the research methodology, review the data collected, and suggest possible alternatives to the experimental design to overcome the problem areas encountered.

ABSTRACT

A research experiment was conducted to investigate how accurate a voice recognition system must be for daily production use. Specifically, the purpose of the research was to establish the percentage accuracy level at which a user becomes frustrated and decides not to use a voice recognition device. The experiment consisted of controlling the perceived recognition accuracy of a voice recognition system and then collecting data through the use of a questionnaire from the experimental users on the acceptability of the equipment. The experiment was not totally successful for a variety of reasons. This paper will discuss the research methodology, review the data collected, and suggest possible alternatives to the experimental design to overcome the problem areas encountered.

User Accuracy Requirements
For Voice Recognition

I. INTRODUCTION

Within the last ten years voice recognition technology has grown from a laboratory research endeavor to a useful and economic, computer human interface tool. The equipment available today is relatively inexpensive, compact, and accurate as evidenced by numerous applications, both industrial and military, which successfully use this input methodology. Improvements in the technology are still being made. Three major areas for continued research in voice input to computers are as follows.

1. User independence
2. Continuous speech
3. Vocabulary capability of 10,000 to 12,000 words

The literature indicates that the solution or at least incremental breakthroughs to these three research areas are just a few years away. Hopefully, the goal of manufacturers and researchers is to develop and produce systems with all or some of the properties listed above which have the best accuracy rate possible. A question arises with this goal in mind. At what accuracy rate should the system be made available to users? If an accuracy rate of 95% is acceptable why not use those systems while new and better algorithms are being developed? Conversly, if 95% accuracy rate is not acceptable and will give voice recognition a bad reputation among management personnel and users it should be held back until an acceptable rate can be achieved. A measure of where this accuracy threshold is could be of

great use to researchers and system manufacturers. Therefore, the question investigated was how poor must a voice recognition system be before the user becomes frustrated with the error rate and will choose not to use the system.

A required accuracy rate will depend on the task for which the system is intended to be used. For example, if voice recognition is to be used for the input of guidance parameters and launch sequence commands in a computerized missile delivery system, anything less than 99+% accuracy would be unacceptable even under the high stress situations most likely existing during the input process. On the other hand, there are numerous tasks which can be labeled as non-critical where voice recognition errors can be tolerated and in fact now occur frequently using more conventional input methods such as a keyboard. A typical task in this category would be information retrieval such as a stock broker obtaining stock information for a client, or an airline reservationist retrieving flight information for a customer. Neither of these tasks are particularly critical in nature. If an occasional error is made during the input process it is easily corrected and the process repeated without any damage to the system or loss in revenue, profit or system efficiency.

An experiment was developed to study the user acceptabiltiy question with the objective of determining a user required accuracy rate. A VET/2 voice recognition unit manufactured by Scott Instrument Inc. of Denton, Texas was used in conjunction with a Basic software program created on an Apple II Plus microcomputer. The experiment's subjects were asked to read an ordered list of words into the voice recognition unit and observe whether the word was recognized properly. This same ordered list of words was stored in

computer memory. Each subject was led to believe the system was in a recognition mode, but in fact, the software program waited for a verbal input and then drew a random number to determine if the spoken input should be displayed as correct recognition, misrecognition or nonrecognition. If the random number determined it was to be a correct recognition the next word on the ordered list was displayed. If a misrecognition was to occur an alternate word was displayed, and a question mark was displayed if the random number determined a nonrecognition was to appear. In other words, the program can be viewed as a voice actuated system where any verbal response would trigger the recognizer, but a random number would determine the output, not the recognition algorithms. As long as the subjects continued to read the ordered list of words, it appeared as if word recognition was being accomplished. By varying the random number test the subject's perceived accuracy rate could be controlled. The errors were recorded, and each subject was asked to complete a questionnaire. The questions were designed to indicate when a frustration level was reached due to recognition inaccuracy.

The experiment did not lead to the desired goals or results. This paper will cover the experiment and why the actual results were different than predicted. First, the report will describe the computer software program which was created to vary the perceived recognition rate during the experiment. Next, the experimental design will be discussed. The entire design was not implemented because preliminary data analysis indicated the desired results were not being obtained. The method used to implement the experimental design will also be discussed. The third section of the report will cover the preliminary data analysis and summary of user responses to the experiment.

Finally, conclusions will be drawn as to why the experiment did not resolve the question at hand, and recommendations will be presented for future research concerning the question of "How good should voice recognition equipment be?".

II. PROGRAM SOFTWARE DESCRIPTION

Two computer software programs were written for use in this experiment. The first program was used to create three databases consisting of words to be used for the planned experimental design. Since this program was used only as an aid for the database preparation, it will not be discussed. A copy of the program and the databases of words are attached in Appendix A and Appendix B respectively. The second program created was used to alter the perceived recognition accuracy of the Scott Instrument VET/2 recognizer. It was written in Applesoft Basic and is included as Appendix C. The following description explains the program logic using the program line numbers for reference.

Line 10 dimensions a character array called W$(200,2). During the experiment, this array held the 200 vocabulary words used for the experiment. W$(I,1) held the word which was to be spoken, and W$(I,2) held a sound alike word. The use of this sound alike word will be explained later. The array F$(3) held the name of the three word databases which were available for use. These databases were named:

1. COMPUTER, designed to be used by users of a text editor on the IBM 3033 computer at the Naval Postgraduate School,
2. STOCK, designed to be used by stockbrokers, and
3. AIRLINE, designed to be used by airline reservationists or travel agents.

The array P(4) held the four different probabilities associated with the planned recognition rates of 99%, 95%, 90%, and 85%.

Line 20 sets the variable D$ to a control D which is used for file manipulation on the Apple computer.

Lines 30 and 40 are data assignment statements for the F$ and P arrays.

Line 45 assigns to the character variable M$ a series of five blank spaces. M$ is used for print control or print spacing.

Line 50 is an output statement asking the user what database is to be used.

Line 60 accepts as an input the number of the desired database. This number is placed in the variable A.

Line 70 opens the correct database file.

Line 80 sets the read device to the appropriate database file.

Line 90 reads in the 200 word vocabulary and their sound alikes into the W$ array.

Line 100 closes the input database file.

Line 105 asks for a random number generator seed and places the integer response into the variable called IS.

Line 110 asks for an algorithm number, which in effect is the array index of the desired accuracy rate. The question in line 110 was stated in such a manner so that experiment participants would assume different word recognition algorithms were being tested.

Line 120 places the algorithm number in the variable B, and if B is less than zero, the program is stopped.

Lines 130 through 160 call the Scott instrument voice recognition subroutines used to load the voice patterns into

memory, and initialize the recognition unit. The voice patterns, although not used for recognition purposes, were neccessary for the proper operation of the recognition unit.

Line 170 prints a header announcing the practice session of 10 words. During this practice session the use of voice recognition equipment was explained to the experiment participant. The explanation given to each participant will be described in detail later.

Lines 180 through 200 create a program loop. The variable I is used as an indexing variable. This indexing variable is first set to 1, and the first word of the vocabulary database is printed on the Apple computer display. This display is used as a prompt to the experiment's participants for the word they are to speak. After the word is displayed the program is transferred to the recognizer's subroutine which will accept a verbal input. After the recognizer accepts the voice input the subroutine returns control to this main program. After the acceptance of the voice response an artificial delay is created by the "FOR Z ..." statement. This delay was neccessary to provide a capability of stopping program execution if the participant made an error which could lead them to believe the program was not actually recognizing their voice. The delay provided the time for the experimenter to stop the program before the "recognized" word was displayed. After the delay a subroutine, which will be described later, is called to determine whether a correct response, a sound alike mistake, a random mistake, or a nonrecognition response should be displayed to the participant. After the response is displayed the index variable, I, is checked to determine if the test practice session is over.

Line 210 sets a series of counters to zero. These counters keep track of the number of non recognitions presented to the participant (Q), the number of sound alike misrecognitions presented (S), and the number of nonsense misrecognition responses (N). The variable T is calculated at the end of each particpant's pass through the 200 word vocabulary list and holds the accuracy rate actually presented to the subject.

Lines 220 through 250 comprise another program loop. The logic is similar to lines 180 through 200 except the entire 200 word vocabulary is sequentially displayed.

Lines 260, 270 and 280 respectively display statements thanking the participants, calculating the actual accuracy rate presented, and displaying the accuracy rate and all counters in a coded form.

Line 290 sends control of the program back to the question asking which algorithm should be used (Line 110), and the program is ready for the next participant.

Line 300 is the first line of the subroutine which will calculate whether the response which is to be shown to the participant is a correct recognition response or one of the three possible error responses. A random number is drawn. If the random number is greater than the accuracy rate which is presently being simulated the program will branch to the statements neccessary to calculate the type of error which should be presented.

Line 310 is executed if a correctly recognized response is to be displayed. The print statement will first print the variable containing the blank spaces, and then the correctly recognized word. This was done so the recognized word was displayed further to the right on the Apple screen

than the word which was output as a prompt. Transfer then is passed to statement 400.

Lines 320 trhough 340 are used to determine the type of error that should be displayed given that the present word is to be perceived as an error by the participant. A random number is drawn, and if it is less than .33 the error is considered a sound alike error and control is passed to line 390. If the random number, R, is between .33 and .66, it is considered a random or nonsense error and control is passed to line 360. Finally, if the random number is greater than .66, it is considered a nonrecognition. Therefore, the three types of errors are equally likely. In previous voice recognition studies, Poock(1980), Poock(1981), and Jay(1981), error rates of about 1.8% have been consistently experienced with a Threshold Technology Inc., Delran New Jersy, model 600 voice recognizer. In these studies non-recognitions consisted of 31% to 35% of the total recorded errors. There were no statistics available on the percentage of misrecognitions which could be considered sound alikes or non-intuitive confusing misrecognitions. For this reason it was assumed that misrecognitions should be equally divided between the sound alike possibilities and the random error possibilities. Therefore, all three error types were programmed to occur with equally likely probabilities.

Line 350 is executed if a non recognition is to occur. It again prints the variable, M$, which contains blanks and then a question mark, ?, representing the Scott Instrument convention for a non-recognition. The counter for non-recognitions is increased by one and the program is then transferred to line 400.

Line 360 through 380 determine a random word response for

the random misrecognition case. A random number is drawn and converted into a random integer between 1 and 200. Next it is checked to ascertain that the random integer generated is not the word which is to be misrecognized. If this check had not been made, a misrecognition could have been recorded but the participant would have in fact seen the correct response. The randomly selected word is printed on the display in the same manner as a properly recognized word, and the nonsense word counter is incremented. Again control is then passed to statement 400.

Line 390 is executed if the random number, R, indicates that this incorrect recognition should be a sound alike and the second word in the W$ array is printed out. These words have been selected in such away that an average user would conclude that it was an understandable or likely recognition error. The sound alike counter is incremented.

Line 400 is the last line of the subroutine. Again a loop is added to produce a delay. The index used in the program loop consisting of statements 220 through 260 is incremented. A blank line is displayed for readablity and the subroutine returns to the main program to print the next word in the 200 word vocabulary list.

The next section will describe how this program was used, and the reaction of the experiment's participants as to the believability that a voice recognizer was being tested.

III. EXPERIMENTAL DESIGN

It has already been mentioned that three databases were formed for the research experiment. The plan was to run three groups of 100 people each through one of the databases. In other words, 100 stockbrokers or investment counselors would say the 200 words associated with their profession, and then rate the acceptability of the equipment. Likewise, 100 airline reservationists or travel agents were to use the airline reservationist word list, and 100 students would use the 200 words associated with the IBM 3033 text editor program.

A word list of 200 words was used for two reasons. First it was decided that going through a list of 100 words went too fast, and the subjects would not get a good feeling for the accuracy rate. Using 300 words was definitely out of the question because of the time involved in conducting the experiment for the number of planned subjects, and because of a possible *boredom factor* which could complicate the subjects perception of the system. The median seemed like a reasonable choice. The vocabulary size of 200 words also satisfied a second criteria. That was the desire to get user frustration information at a more accurate level than every 1 percentage point. At least the 200 words would give ratings at every one half percentage point.

During the experiment it became apparent that a frustration level was not being achieved or at least measured. Therefore, after the first fourth of the experiment some preliminary data analysis was done, and the results showed that little was being learned about user accuracy needs. It was decided to stop the experiment and report on what had been done to date. A total of 78

subjects participated in the test. The intent of this section of the paper is to explain how the subjects were introduced to the experiment and how they reacted to the recognition system. This section will also discuss the design of the questionnaire.

The subjects were students, staff and faculty members at The Naval Postgraduate School in Monterey, California. They were all volunteers and between the ages of 25 and 49. The entire explanation and experiment took between 10 and 15 minutes per subject.

When a subject arrived at the experiment site, it was explained that some new user independent voice recognition algorithms were being tested. The Scott Instrument voice recognition device was covered to preclude the subjects from getting the wrong impression of its capability. The algorithm number was entered into the system in front of the subject as the idea of different algorithms was being explained.

Some of the subjects had used voice recognition equipment on previous voice experiments. These students needed little practice, but still went through the ten practice words. The practice words were used as the teaching device for those subjects who had not used the equipment before. The need to speak a phrase as a continuous flow of speech was explained, as was the explanation of the meaning of a question mark (?) when it appeared on the display. If the subject showed an interest in the machine's capability during the the practice session, their questions were postponed until after they had answered the questionnaire.

The subjects were asked to ignore all the errors (if any) which occured during their practice session. It was explained that some artificial intelligence algorithms were

being employed and the system was selecting characteristics of their voice for use in the main portion of the experiment.

As the subjects were going through the 200 word list, the experiment was stopped each time an error occured. The error was pointed out to the subject and a short explanation was given as to what that error would do if the true text editor was employed. The idea for stopping the program was two fold. First it was noticed during preliminary program testing that some people started to read the words on the display and weren't watching the recognized word displayed. In other words, the experiment bored them and they weren't always aware of the errors. At first this had been solved by placing the recognized word directly under the prompt word. Unfortunately this solution caused another problem. The subjects who participated in the preliminary testing started to say the recognized word instead of the following prompt word which had a devastating effect on the user's confidence that the system was recognizing their voice. Therefore, the recognized word was placed to the right of the prompt word, and each error was pointed out to the subject. The second reason for stopping the program was to delay the subject in completing the experiment. It was hoped that the idea of an error slowing them down would transfer to their perception of how the system would work in a real environment. The plan was to slow them down thus creating a frustration level which was to be measured.

There were numerous times even with the precautions taken where the subject read the wrong word, or started to make a comment without the microphone being turned off. This led to a recognition which in the majority of cases was correct, when the spoken utterance was obviously incorrect. The experimenter's finger was always kept lightly on the Apple

computer's space bar. By depressing the bar the microphone and voice recognition system were deactivated. The majority of the time, the experimenter depressed the space bar soon enough to avoid a correct recognition of an incorrect voice input. This was the major reason for the delay loop explained earlier in the program software description. If the space bar was not depressed soon enough, and the correct response appeared on the display, the subject was told that the software program had been developed to do its own data collection automatically. Furthermore, it was explained that the experiment was interested in only recognizer errors not the human errors which will always occur with a voice system. Therefore, it was explained, that pressing the space bar was an automatic override, and no matter what was said the automatic data collection routine would count it as a correct recognition. This explanation seemed to satisfy everyone, who encountered the situation.

After the subject had completed the 200 words, he or she was asked to fill out a two page questionnaire. This questionnaire is attached as Appendix D. Questions 3,4 and 5 created the data which was of most importance to the experiment and was the measurement of user acceptability. The sets of response alternatives, for these three questions, were taken from an Army Research Institute publication on questionnaire construction (1976). The responses have been tested and shown to have mean scaling factors at least one standard deviation away from each other while maintaining the parallel wording. Question 9, concerning the part of the country the subject grew up in, had nothing to do with the experiment that was being conducted, but was included to make the experiment about testing user independent algorithms more believable to the participants. The remaining questions are self explanatory,

and will be reviewed in detail when discussing the analysis of the data collected.

After the questionnaire was completed the subject was free to ask questions about the system. They were led to believe that all algorithms had been created at the School, and that the technology was not commercially ready because of the extremely large amount of core needed to run the system. This was done to insure that the believability in the system would not decrease as students talked about the experiment. On the other hand, it was not the intent of the experiment to lead the subjects to believe that voice recognition capabilities were beyond the present state of the art.

The experiment was a total success as far as the believability of the system was concerned. There were a couple of instances when the random number generator cooperated fully. For example, a subject asked a question which triggered a response and the random number generator created a nonrecognition. In another case a word was misrecognized in the test sequence and the same word was misrecognized during the experiment, both times the misrecognition was the sound alike word. There were subjects who tried to analyze the system and hypothesized why the system did not recognize them correctly. One subject was convinced that any word with an "S" sound would not be recognized properly because he tended to slur his "S" sound. There were only two subjects out of the 78 tested who mentioned the fact that they doubted the system was recognizing their voice. This fact was noted on their questionnaire after they left the laboratory area.

IV. DATA ANALYSIS

Appendix E and F present the data collected. Appendix E contains the raw data collected, while Appendix F has the data in the ranked form. All of the data analysis used nonparametric statistics methodologies based on ranks.

The first column of data in Appendix E contains the total number of errors the user observed. This number is the sum of nonrecognitions and both types of misrecognitions. The second column contains the total number of misrecognitions which is the sum of the sound alike errors and nonsense errors which were observed by the subject. Columns 3,4, and 5 are the individual error totals for nonrecognitions, sound alike misrecognnitions, and random misrecognitions, respectively. Column 6 is the age of the subject, and column 7 is the average number of hours the subject spends using a computer terminal each day. The average number of hours spent at a terminal were considered important for the stock brokers and airline reservationists, but had little meaning for the students who became the only participants in the experiment. Therefore, this data will not be used in the analysis, but is presented for completeness.

The next two columns, column 8 and 9, are the answers to the questions on whether the subject had ever seen or used voice recognition equipment. These answers are coded with a 1 representing an answer of "yes", and a 2 representing a "no" response. Columns 10, 11, and 12 are the responses to the questions dealing with user acceptability. These are also coded, from a 1 meaning a poor acceptability response to 5 for a high acceptability response. Column 13 is the sum of the responses to the three questions. Since the questions each had parallel wording the sum of the answers

for the individual questions was used for the data analysis. This method gave a more accurate numeration of the acceptability level for each subject.

Column 14 and 15 are the data collected on the subject's perception of his or her own typing ability in terms of speed and accuracy respectively. These data fields are also numeric codes for the response given on the questionnaire. A value of 1 represents a poor rating, a 2 an average rating while a 3 represents a very good rating in the speed and accuracy capability of each subject. Column 16 is the response to the question of whether the subject ever had a typing course. The same response convention was used for the two previous yes-no questions.

The second to the last column, column 17, tabulates the data collected representing the geographic region where the subject grew up. The codes have the following meaning.

1 - South
2 - East
3 - Midwest
4 - Foreign
5 - West
6 - All over or not specified

Finally the last column, column 18, is the subject's number of years experience with computers and computer terminals.

As it was already mentioned Appendix F contains the ranked data. The data are organized in the same manner as previously described for the raw data except some of the columns are the ranks of the data collected. There were numerous ties and the rank value assigned was the average of the ranks that would have been assigned to them had there been no ties. All of the recognition error counts were

ranked, as were the age data, total acceptability rating data, and the subject experience data.

The first set of analysis used Spearman's rank correlation coefficient (Conover, 1980) as a test statistic to determine if there was any negative correlation between the number of errors presented to a subject and that subject's acceptability rating. It had been hypothesized that as the number of errors increased the user acceptability would decrease. Therefore, the null hypothesis was that the number of errors and user acceptability are mutually independent or had no correlation. The alternative hypothesis was that these two variables are negatively correlated. The following tabulates Spearman's correlation coefficient calculated on the ranked data between these two variables.

all errors	-.250
total misrecognitions	-.249
non recognitions	-.250
sound alikes	-.215
random misrecognitions	-.201

At a significance level of .025 the null hypothesis can be rejected for the first three values. It appears as if the desired negative correlation is present, but the correlation is very slight as indicated by the value of the correlation coefficient. The hypothesis of mutual independence can not be rejected between the sound alike errors and user acceptability, and the random errors and the acceptability variable.

Figure 1 is a graph of the numeric user acceptability totals versus total user perceived errors. It shows that although there is evidence of a negative correlation there is little information existing as to where or at what error

level the acceptability values start to decrease. In fact the real problem is exemplified by observing the average response values for various groups of subjects. Table 1 contains, for various groups of subjects, the averages of the total values for the three acceptability questions along with the standard deviation. There is very little difference between the groups. Even the group that observed more than a 15% error rate still rated the system in the "like it" and "would use it" category. A Kruskal-Wallis test was done to determine if these groups of subjects had the identical mean response values. This hypothesis could not be rejected at the .05 or .1 significance level. The test statistic value, T, was calculated at 6.78 and the chi-square distribution quantile for the four degrees of freedom at the .05 level is 9.488 and at the .1 level is 7.779. Therefore, even though a small negative correlation is detected in the data, very little information can be gained as to where a distinct drop occurs in user acceptability values.

Two other Spearman's Rho correlation coefficient tests were done. First a correlation possibility was investigated between the ranked values of age and the ranked values of acceptability. The hypothesis that age and acceptability values were not correlated could not be rejected. Spearman's correlation coefficient was calculated at .03. The same test was done to check the data for mutual independence among the ranked values for years of experience and the ranked values of the acceptability totals. Again the independence hypothesis could not be rejected with a correlation coefficient of -.02.

In addition to the Kruskal-Wallis test previously mentioned, a series of similar tests were completed in order to determine if there were any differences in the mean

acceptability responses among different groups of individuals. Table 2 summarizes these tests. None of the hypotheses that all the mean acceptability responses were idenitical could be rejected. In other words no statisical differences could be found among the various groups tested.

The only test which suggested a possible difference was the test between the groups divided by geography. This was interesting because numerous people approached the experiment asking for example, if this machine understood "Louisianian". They knew they had a distinct southern accent, and if it recognized them they were very surprised. This could account for the relatively high Kruskal-Wallis statistic even though the groups could not statistically be shown to have different means.

In conclusion, there is really very little information present in the data collected. For this reason the experiment was cut short before the time was spent at business establishments. It is hypothesized that there are at least two basic flaws in the experimental design. First the experiment was working with an advanced technology. Some of the people who were tested were not aware that voice recognition existed. In fact, numerous people answered on the questionnaire that they had seen voice recognition to computers on "Star-Trek". With an attitude like that any recognition capability was impressive. If this is true it would be expected that there might have been a larger difference between the group acceptability averages between those who had seen voice recognition equipment before and those who had not. Since this did not occur we can only assume that the group which had seen voice recognition before knew about its user dependence limitations and were equally impressed with the user independence capabilities being demonstrated.

The second problem area involves the lack of a task which needed to be accomplished within the experimental framework. Reading a list of words and pointing out the errors did not create the frustrating situations which are going to exist when you encounter a recognition error while trying to accomplish a task. Although it was hoped that stopping the experiment each time an error occurred would provide this feeling, it did not totally simulate the frustration associated with task completion. Furthermore, subjects who had never seen or used voice recognition equipment had a difficult time visualizing how the equipment would actually be used. Although an explanation was given to each subject at the beginning of the experiment, it appears as if the concept was not totally understood by everyone. This was evidenced by some of the questions asked by the subjects after the experiment was finished. If there had been a realistic job or goal, this problem could be alleviated.

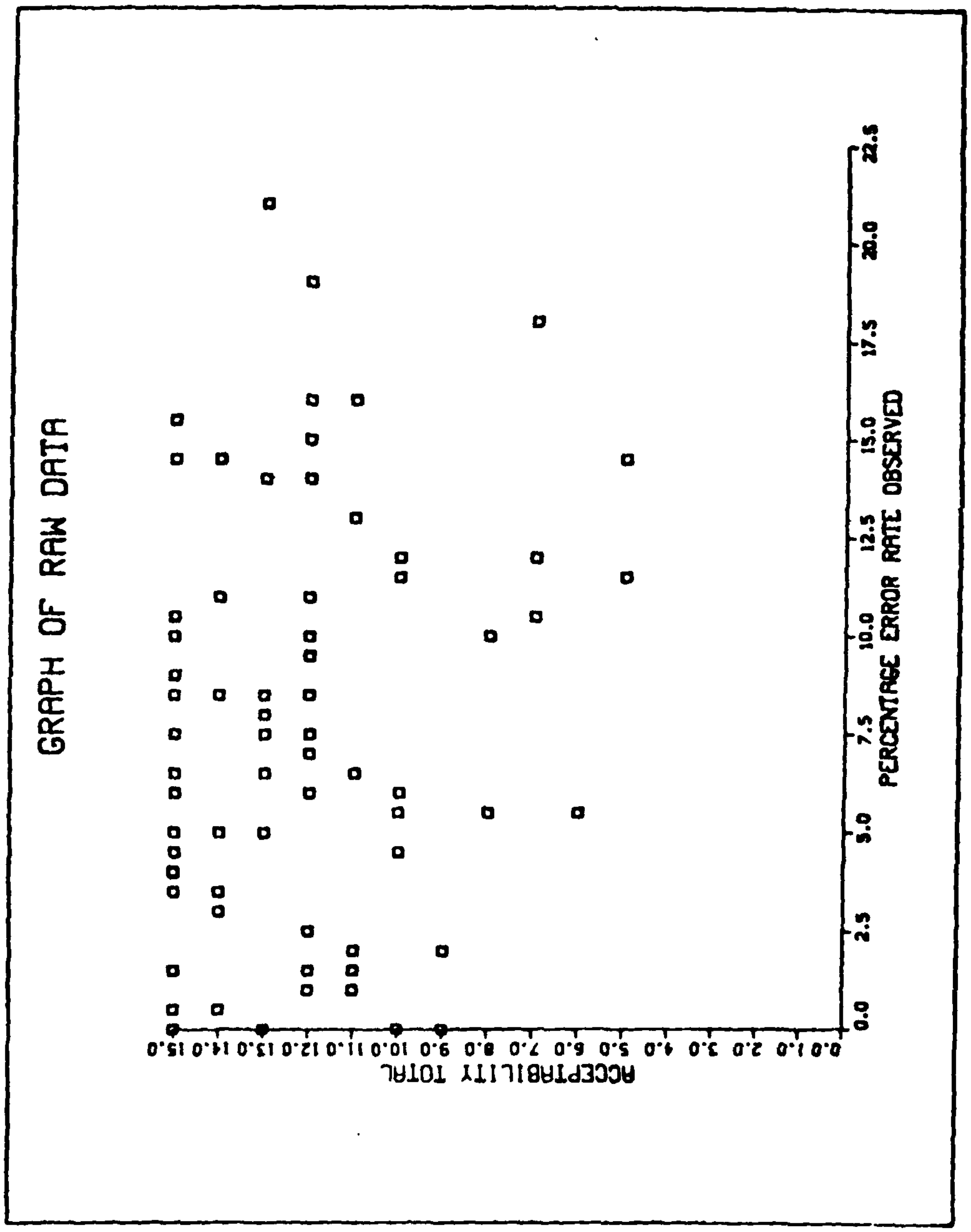


Figure 1

Graph of Raw Data

% Error Rate Groups	# of subjects in group	Average	Standard Deviation
0.0 - 1.5	19	4.40	4.46
2.0 - 5.0	13	4.41	4.49
5.5 - 10.0	23	3.96	4.08
10.5 - 14.5	16	3.71	3.93
15.0 - 21.0	7	3.90	4.02

TABLE 1.

Acceptability Response Summary

GROUP	TEST STATISTIC	CHI-SQUARE = .05
Seen Voice equipment- not seen voice equipment	1.18	3.84
Used Voice equipment- not used voice equipment	3.65	3.84
Slow, intermediate, fast speed typist	2.14	5.99
Poor, fair, very good typing accuracy	4.38	5.99
Typing course- no typing course	0.12	3.84
Part of country raised in	10.15	11.07

TABLE 2.

Results of Kruskal-Wallis Tests

V. RECOMMENDATIONS AND CONCLUSIONS

The data collected in this experiment do not help answer the question of how good a voice recognition system must be to maintain user acceptability. The following recommendations are provided for future research efforts and experiments.

1. Introduce the subjects to voice recognition equipment before the start of the experiment.

2. Demonstrate the equipment within an appropriate work environment. In other words, demonstrate the use of the equipment to accomplish a suitable job within their individual areas of expertise.

3. Have the subjects train the equipment. Don't present a system which is obviously years ahead of currently available technology.

4. Set up a series of tasks which are suitable for voice recognition input which can be accomplished within the test environment. It appears that it is important that the job is very realistic in nature.

5. Consider the following method to vary the recognition accuracy for the experiment. Use a recognition system which has the capability to easily access not only the word which was recognized, but the runner up word. Then by the use of a random process determine whether the recognition unit should output the first or runner up word. If the recognition system used has a fairly good accuracy rate, the the first choice word should be the correct word. Therefore, by

randomly selecting the second word you are randomly selecting errors from the recognition unit. It is possible that the recognition unit will make a mistake and the recognized word will be incorrect. If the random draw determines a correct response is to be given, the first word will be output to the system but in this case it is an error. This error was not expected; therefore, although the actual accuracy rate of the system will not be under the experimenters control the overall error rate should be very close to the percentage of times the runner up word is chosen. If a 100% accurate system was used this percentage of runner up choices would be equal to the error rate observed. Since a 100% accurate system does not exist it appears as if close will have to do.

The Interstate Electronics Corporation machine is suitable for this type of task. The only problem involved with this recommendation is that it will be impossible to observe error rates much less than the underlying error rate associated with the equipment chosen for use in the experiment.

Since there was evidence of the hypothesized negative correlation, it is possible that the frustration measurement will fulfill the needs of follow on experiments. This question of frustration measurement should be investigated further before undertaking the next phase of experimentation to answer the question about acceptable accuracy rates for voice recognition equipment.

BIBLIOGRAPHY

Conover, W. J., Practical Nonparametric Statistics, John Wiley & Sons, 1980.

Jay, G. T., An Experiment in Voice Data Entry for Imagery Intelligence Reporting, Master's Thesis, Naval Postgraduate School, Monterey, 1981.

Nystrom, C. O., Questionnaire Construction Manual, U.S. Army Research Institute for the Behavioral and Social Sciences, 1976.

Poock, G. K., Experiments With Voice Input for Command and Control: Using Voice Input to Operate a Distributed Computer Network, Naval Postgraduate School Report NPS55-80-016, April 1980.

Poock, G. K., A Longitudinal Study of Computer Voice Recognition Performance and Vocabulary Size, Naval Postgraduate School Report NPS55-81-013, June 1981.

Scott Instrument Inc., VET/2 Voice Entry Terminal Operator's Manual, Scott Instrument, Denton, Texas, 1981.

```
+LIST 1,9000

10 D$ = "": REM  D$ CONTAINS A CNTRL D
20 I$ = ": REM  I$ CONTAINS A CNTRL I
30  DIM F$(3):F$(1) = "COMPUTER":F$(2) = "STOCK":F$(3) = "AIRLINE"
40  DIM W$(200,2)
50  PRINT "DO YOU WANT TO:": PRINT "1. CREATE A NEW FILE": PRINT "2. UPDAT
     E A FILE": PRINT "3. PRINTOUT A FILE": PRINT "4. STOP"
60  INPUT A
70  ON A GOTO 1000,2000,3000,4000
1000 B$ = "CREATE"
1010  GOSUB 6000
1020  PRINT D$;"OPEN ";F$(A);",V001": PRINT D$;"CLOSE"
1030 I = 1:J = 1
1040  GOSUB 5000
1050  GOTO 50
2000 B$ = "UPDATE"
2010  GOSUB 6000
2020  ONERR  GOTO 2060
2030  PRINT D$;"OPEN ";F$(A);",V001"
2040  PRINT D$;"READ ";F$(A)
2050  FOR I = 1 TO 200: FOR J = 1 TO 2: INPUT W$(I,J): NEXT J: NEXT I
2060  PRINT D$;"CLOSE"
2070  PRINT "DO YOU WANT TO:": PRINT "1. CHANGE INDIVIDUAL WORDS": PRINT "
     2. ADD TO THE LIST OF WORDS": PRINT "3. STOP"
2080  INPUT AA
2090  ON AA GOTO 2100,2200,2300
2100  PRINT "WHAT WORD NUMBER DO YOU WANT TO CHANGE": PRINT "IF YOU WANT T
     O STOP INPUT A -1"
2110  INPUT I: IF I < 0 THEN  GOTO 2170
2120  PRINT "DO YOU WANT TO CHANGE": PRINT "1. THE WORD": PRINT "2. THE SO
     UND ALIKE"
2130  INPUT J
2140  PRINT I,J
2150  INPUT W$(I,J)
2160  GOTO 2100
2170 I = 201: GOSUB 5070
2180  GOTO 2070
2200  PRINT "WHAT WORD NUMBER DO YOU WANT TO START AT?"
2210  INPUT I:J = 1
2220  GOSUB 5000
2230  GOTO 2070
2300  GOTO 50
3000 B$ = "PRINTOUT"
3010  GOSUB 6000
3020  ONERR  GOTO 3060
3030  PRINT D$;"OPEN ";F$(A);",V001"
3040  PRINT D$;"READ ";F$(A)
3050  FOR I = 1 TO 200: FOR J = 1 TO 2: INPUT W$(I,J): NEXT J: NEXT I
3060  PRINT D$;"CLOSE"
3070  PRINT "DO YOU WANT": PRINT "1. JUST THE WORD LIST": PRINT "2. WORD L
     IST AND SOUND ALIKES": PRINT "3. STOP"
3080  INPUT J
3085  ON J GOTO 3090,3090,3400
```

```
3090  PRINT "DO YOU HAVE OKI-IMAGE LOADED AND WANT A HARDCOPY? (Y OR N)"
3100  INPUT Y$
3110  IF Y$ = "N" THEN  GOTO 3170
3120  PRINT D$;"PR# 2"
3130  PRINT I$,"80N"
3170  ON J GOTO 3200,3300,3400
3200  FOR I = 1 TO 200: PRINT I,W$(I,J): NEXT I
3210  PRINT D$;"PR# 0": GOTO 3070
3300  FOR I = 1 TO 200: PRINT I,W$(I,J - 1),W$(I,J): NEXT I
3310  PRINT D$;"PR# 0": GOTO 3070
3400  GOTO 50
4000  STOP : END
5000  IF I ) 200 THEN  GOTO 5065
5010  PRINT I,J
5020  INPUT W$(I,J)
5030  IF W$(I,J) = "END" THEN  GOTO 5065
5040 J = J + 1
5050  IF J = 2 THEN  GOTO 5010
5060 I = I + 1:J = 1: GOTO 5000
5065  GOSUB 5070: RETURN
5070  PRINT D$;"OPEN ";F$(A);",V001"
5080  PRINT D$;"WRITE ";F$(A)
5090  FOR K = 1 TO I - 1: FOR J = 1 TO 2: PRINT W$(K,J): NEXT J: NEXT K
5100  PRINT D$;"CLOSE": RETURN
6000  PRINT "WHAT FILE DO YOU WANT TO ";B$: PRINT "1. COMPUTER": PRINT "2.
      STOCK": PRINT "3. AIRLINE"
6010  INPUT A: RETURN
```

Airline Vocabulary

1.	Allentown	Morgantown
2.	Abilene	Aberdeen
3.	Albuquerque	Kodiak
4.	East Hampton	Birmingham
5.	Washington Dulles	Washington National
6.	Columbus Ohio	Columbus, Georgia
7.	Akron	Charleston
8.	Ohara Field	Omaha
9.	Cleveland	Lakeland
10.	Cedar Rapids	Grand Rapids
11.	Corpus Christi	Cincinnati
12.	San Juan	South Bend
13.	Boston	Boise
14.	Rapid City	Sioux City
15.	Daytona Beach	Long Beach
16.	Goose Bay	Pine Coy
17.	Dayton	Galveston
18.	Utica	Sitka
19.	Denver	Beaver
20.	Lewiston	Allentown
21.	Detroit	Duluth
22.	Montpelier	Eau Claire
23.	Charleston	Kinston
24.	Monterey	San Jose
25.	Franklin	Richland
26.	Bangor	Boston
27.	Fargo	Elko
28.	Manitowoc	Roanoke
29.	New York	Norfolk
30.	San Jose	San Diego
31.	Fort Collins	Ft Smith
32.	Laramie	Miami
33.	Kona	Helo
34.	San Diego	San Francisco
35.	Freeport	Bridgeport
36.	Missoula	Honolulu
37.	Oklahoma City	Jefferson City
38.	Colorado Springs	Hot Springs
39.	Pittsburgh	Plattsburgh
40.	Duluth	Detroit
41.	Lousville	Knoxville
42.	St Petersburg	Rochford
43.	Honolulu	Missoula
44.	Kirksville	Louisville
45.	Columbia	Atlanta
46.	Columbus Georgia	Columbus, Ohio
47.	Roanoke	Manitowoc
48.	Rockford	Medford
49.	Washington National	Washington Dulles
50.	Jacksonville	Jackson

51. Atlantic City	Oklahmoa City
52. Chico	Toledo
53. Amaha	OHara Field
54. Plattsburgh	Pittsburgh
55. Eau Claire	Montpelier
56. Norfolk	New York
57. Cheyenne	San Juan
58. San Jose	Monterey
59. Phellipsburgh	St. Petersburg
60. Helo	Kona
61. Medford	New Bedford
62. Beaver	Denver
63. Lewistown	Lewiston
64. Sitka	Utica
65. Galveston	Dayton
66. Grand Rapids	Cedar Rapids
67. Dubuque	Albuquerque
68. Pine Coy	Goose Bay
69. Richland	Cleveland
70. Long Beach	Daytona Beach
71. South Bend	Cheyenne
72. Birmingham	East Hampton
73. Sioux City	Rapid City
74. Cincinnati	Corpus Christi
75. Kenston	Akron
76. Acapulco	Chico
77. Nome	Rome
78. Orlando	Kokoma
79. Pompano Beach	West Palm Beach
80. Harrisburg	Paris
81. Hot Springs	Colorado Springs
82. Pittsfield ,	Springfield
83. Kodiak	Dubuque
84. Milwaukee	Menominee
85. Knoxvill	Kirksville
86. Ft. Smith	Fort Collins
87. Paris	Harrisburg
88. Morgantown	Lewistown
89. Springfield	Pittsfield
90. Elko	Fargo
91. New Bedford	Phillipsburg
92. West Palm Beach	Pompana Beach
93. Bridgeport	Freeport
94. Kokomo	Orlando
95. Aberdeen	Abilene
96. Jackson	Jacksonville
97. Menomimee	Milwaukee
98. Fort Wayne	Fort Lauderdale
99. Rome	Nome
100. Toledo	Acopulco
101. Rochester	Worchester
102. Kalamozoo	Waterloo
103. Providence	Provincetown

104. Morristown
105. Clarksburg
106. Tallahassee
107. San Angelo
108. Martha Vineyard
109. Gainesville
110. Sacremento
111. Lake Charles
112. Osage Beach
113. Brookings
114. Pueblu
115. Antiguo
116. Knoxville
117. Larado
118. Ogdensburg
119. Sumter
120. Show Low
121. Wichita Falls
122. Worthington
123. Provincetown
124. Waterloo
125. Apple Valley
126. Appleton
127. Modesto
128. Worchester
129. Huntsville
130. Waterville
131. Baton Rouge
132. Marquette
133. New Orleans
134. Walla Walla
135. Tupelo
136. Astoria
137. Catskills
138. Atlanta
139. Janesville
140. Durango
141. Newburg
142. Trenton
143. Billings
144. Fayetteville
145. Greenvill
146. Buffalo
147. Evansville
148. Phoenix
149. Greensboro
150. Grand Junction
151. Anchorage
152. Wichita Falls
153. San Francisco
154. Waterville
155. San Juan
156. Princeton

Worthington
Marthas Vineyard
Apple Valley
Show Low
Ogdensburg
Knoxville
Larado
Lake Tahoe
Anchorage
New Orleans
San Angelo
Altoona
Waterville
El Dorado
Lynchburg
Rochester
Tampico
Wichita
Riverton
Providence
Kalamazoo
Tallahassee
Trenton
Durango
Sumter
Catskills
Janesville
Osage Beach
Phoenix
Billings
Antingua
Modesto
Augusta
Greinville
Bermuda
Huntsville
Buffalo
Clarksburg
Grand Junction
Brookings
Gainsesville
Evansville
Greensboro
Jacksonville
Marquette
Ontario
Morristown
Baton Rouge
Twin Falls
San Juan
Greenville
San Jose
Pendleton

157.	Churchill Falls	Idaho Falls
158.	Hastings	Harrison
159.	Wichita	Wichita Falls
160.	Fort Lauderdale	Fort Wayne
161.	Pueblo	Sacramento
162.	Saint Louis	St. Croix
163.	El Dorado	Pueblo
164.	Grand Cayman	Grand Canyon
165.	Jefferson City	Atlantic City
166.	Idaho Falls	Churchill Falls
167.	Pendleton	Princeton
168.	Greenville	Waterville
169.	Miami	Laramie
170.	Grand Canyon	Grand Cayman
171.	St Croix	Saint Louis
172.	Lakeland	Franklin
173.	Harrison	Hastings
174.	Boise	Bangor
175.	Twin Falls	Wichita Falls
176.	Riverton	Clinton
177.	Ontario	Orlando
178.	Lynchburg	Galesburg
179.	Altoona	Astoria
180.	Agusta	Wausau
181.	Victoria	Topeka
182.	Sarosota	Pensacola
183.	Bermuda	Columbia
184.	Poplar Bluff	Council Bluffs
185.	Tulsa	Guadalahara
186.	Jacksonville	Danville
187.	Poza Rica	Walla Walla
188.	Guadalahara	Sarasota
189.	Pensacola	Poza Rica
190.	Orlando	Mankato
191.	Topeka	Tillsa
192.	Wausau	Victoria
193.	Galesburg	Newburg
194.	Penta Gorda	Tupelo
195.	Lake Tahoe	Lake Charles
196.	Danville	Fayetteville
197.	Mankato	Pueblo
198.	Hattiesburg	Clarksburg
199.	Lampico	Punta Gorda
200.	Clinton	Appleton

Computer vocabulary

1	Login	Logoff
2	Alter	Add
3	Backward	Bottom
4	Command	Compress
5	Cursor Column	Cursor File
6	Findup	Forward
7	Left	Load
8	Nfind	Nfind-up
9	Put-D	Parse
10	Query	Quit
11	Reset	Restore
12	Set Autosave	Set Case
13	Set File Mode	Set File Name
14	Set Line Character Off	Set Logical Record Length
15	Set Number	Set Pack
16	Set Reserved	Set Scale
17	Set Synonym	Set Tableine
18	Set Verify	Set Wrap
19	Stack	Status
20	Delete Line	Duplicate Line
21	Assign	Access
22	Five	Four
23	Six	Sixty
24	Seven	Seventy
25	Eight	Eighty
26	Nine	Ninety
27	Ten	Send
28	Alpha	Papa
29	Bravo	Romeo
30	Charlie	Whiskey
31	Logoff	Login
32	Bottom	Backward
33	Cmsg	Command
34	C-Repalce	Cursor Column
35	Find	Findup
36	Join Cursor	Join
37	Msg	Macro
38	Next	Nfind
39	Purge	Put-D
40	Replace	Reset
41	Set Arbitrary Character	Set Autosave
42	Set Filler	Set File Mode
43	Set Line Character On	Set Line Character Off
44	Set Nulls	Set Number
45	Set Record Format	Set Reserved
46	Set Stream	Set Synonym
47	Set Variable Blank	Set Verify
48	Split	Stack
49	Uppercase	Up
50	Duplicate Line	Delete Line
51	Access	Assign
52	Filedef	Fetch
53	Global	Cobol
54	Listfile	Filedef
55	Start	State

56	Zero	Kilo
57	One	Echo
58	Two	Echo
59	Three	Thirty
60	Four	Five
61	Profile	No Profile
62	Cancel	C-Append
63	Cms	Cmsg
64	Cp	C-Replace
65	Expand	Emsg
66	File	Find
67	Join Column	Join Cursor
68	Modify	Move
69	Power-Input	Preserve
70	Renumber	Repeat
71	Selective Change	Set APL
72	Set Current Line	Set Escape
73	Set Image	Set Implicityly to CMS
74	Set Message Mode	Set Nondisplayable Characters
75	Set Prefix	Set Range
76	Set Span	Set Stay
77	Set Text	Set Truncate Column
78	Sort	Sos
79	Type	Up
80	Debug	Disk
81	Papa	Alpha
82	Quebec	Stack
83	Romeo	Tango
84	Sierra	Alpha
85	Tango	Romeo
86	Uniform	Nine
87	Victor	Preserve
88	Whiskey	X-ray
89	Yankee	Charlie
90	Zulu	Move
91	No Profile	Profile
92	C-Append	Cancel
93	C-Locate	Cms
94	C-overlay	Cp
95	Duplicate	Down
96	Emsg	Expand
97	Put	Input
98	Join	Join Column
99	Move	Msg
100	Preserve	Purge
101	Repeat	Replace
102	Set APL	Set Arbitrary Character
103	Set Escape	Seet Filler
104	Set Implicitly Cms	Set Line Character On
105	Set Nondisplayable Characters	Set Nulls
106	Set Range	Set Record Format
107	Set Stay	Set Stream
108	Set Truncate Column	Set Variable Blank
109	Sos	Split
110	Up	Uppercase

111	Disk	Debug
112	X-Ray	Yankee
113	Twenty	Ninety
114	Thirty	Eighty
115	Forty	Seventy
116	Fifty	Five
117	Sixty	Fifty
118	Seventy	Seven
119	Eighty	Thirty
120	Ninety	Twenty
121	Xedit	Update
122	C-Delete	C-First
123	C-Last	C-Locate
124	Compress	Copy
125	Cursor File	Cursor Screen
126	Forward	Find-up
127	Get	Set
128	Load	Locate
129	Nfind-up	Next
130	Overlay	On
131	Quit	Q-Quit
132	Restore	Right
133	Set Case	Set Command Line
134	Set File Name	Set File Type
135	Set Logical Word Length	Set Macro
136	Set Pack	Set Program Function Key
137	Set Scale	Set Screen
138	Set Tabline	Set Tabls
139	Set Wrap	Set Zone
140	Status	Stack
141	Top	Transfer
142	Compare	Cp
143	Erase	Disk
144	Fetch	Filedef
145	State	Start
146	Delta	India
147	Echo	One
148	Foxtrot	Findup
149	Gulf	Off
150	Hotel	Help
151	Update	Xedit
152	C-First	Change
153	C-Insert	C-Last
154	Copy	Count
155	Cursor Screen	Cursor File
156	Delete	Down
157	Set	Get
158	Help	Hextype
159	Locate	Lowercase
160	On	Off
161	Q-Quit	Query
162	Read	Recover
163	Right	Read
164	Set Command Line	Set Column Pointer
165	Set File Type	Set Hexidecimal

166	Set Macro	Set Mask
167	Set Program Function Key	Set Point
168	Set Screen	Set Serial
169	Set Tabls	Set Terminal
170	Set Zone	Set Equals
171	Transfer	Type
172	CP	Compare
173	Cobol	Global
174	India	Lima
175	Juliett	Delete
176	Kilo	Zero
177	Lima	India
178	Mike	Quit
179	November	Oscar
180	Oscar	Alter
181	Add	Alter
182	Change	C-Insert
183	Count	C-Overlay
184	Down	Duplicate
185	Hextype	Help
186	Input	Put
187	Lowercase	Left
188	Macro	Modify
189	Off	Overlay
190	Parse	Power-Input
191	Recover	Renumber
192	Save	Selective Change
193	Set Column Pointer	Set Current Line
194	Set Hexidecimal	Set Image
195	Set Mask	Set Message Mode
196	Set Point	Set Prefix
197	Set Serial	Set Span
198	Set Terminal	Set Text
199	Set Equals	Set Prefix
200	Shift	Sort

Stock vocabulary

1	Houston Natural Gas	Oklahoma Gas
2	IBM	MGM
3	ITT	IBM
4	Iowa Electric	Iowa Gas
5	K-Mart	Gillette
6	Kaiser Aluminum	Kaiser Steel
7	Lenox	Purex
8	Leer-Siegler	Singer
9	Lionel	Mattel
10	Iowa Gas	Iowa Electric
11	Litton	Haliburton
12	Kaiser Steel	Kaiser Aluminum
13	Lockheed	Square D
14	MGM	IBM
15	Magic Chef	Zenith
16	Marriott	Mattel
17	Marathan Oil	Ashland Oil
18	Mattel	Marriott
19	Maytag	Rubber Maid
20	Memorex	Tampox
21	McDonalds	McDonald-Douglas
22	AMF	ATT
23	McDonald-Douglas	McDonalds
24	3M	BDM
25	NCR	CBS
26	Nabisco	Sambos
27	Natomas	Ameroda Hess
28	Oklahoma Gas	Houston Natural Gas
29	Penney's	Macy's
30	Combustion Engineering	Cummins Engine
31	Can Edeson	Western Union
32	Dennys	Dorsey
33	Macys	Penneys
34	Dow Chemical	Dow Jones
35	Elgin	Haliburton
36	Dorsey	Dennys
37	Florida Power & Light	Wisconsin Power & Light
38	Greyhound	Grumman
39	Haliburton	Elgin
40	Kroehler	Burroughs
41	Wisconsin Power & Light	Florida Power & Light
42	Grumman	Greyhound
43	U.S. Air	U.S. Gypsum
44	U.S. Steel	U.S. Air
45	U.S. Life	U.S. Steel
46	Western Union	Con Edison
47	Zapata	Aenith
49	Air Florida	Air Wisconsin
50	Wennebago	Whirlpool
51	Air Wisconsin	Air Florida
52	Atlantic Richfield	Atlantic Research
53	BDM	MGM
54	Tampox	Zerox
55	Alabama Power	Florida Power & Light

56	Bank of Virginia	Bank of America
57	Burroughs	Kroehler
58	CBS	NCR
59	Atlantic Research	Atlantic Richfield
60	Zenith	Zapato
61	Xerox	Lenox
62	Bank of America	Bank of Virginia
63	John Deer	Chrysler
64	Delta Airlines	Eastern Airlines
65	Dow Jones	Dow Chemical
66	Eastern Airlines	Delta Airlines
67	General Motors	General Radio
68	General Steel	Data General
69	General Radio	General Motors
70	Georgia Pacific	Canadian Pacific
71	Goodyear	Goodrich
72	Gillette	K-Mart
73	Gulf Oil	Gulf Western
74	Hicla Mining	Dome Mining
75	Goodrich	Goodyear
76	Heinz	Honda
77	Peirex	Clorox
78	Sambos	Collins Radio
79	Singer	Leer Siegler
80	Square D	Tandy
81	Collins Radio	Texaco
82	Tandy	Square D
83	Texaco	Nabisco
84	United Airlines	Eastern Airlines
85	U.S. Gypsum	U.S. Life
86	Ala Moana	Coca Cola
87	Albertsons	Alles Chalmers
88	Gulf Western	Gulf Oil
89	Dome Mining	Hiclo Mining
90	Ameroda Hess	Natomas
91	American Broadcasting	American Hospital
92	Honda	Heinz
93	Allis Chalmers	Albertsons
94	American Hospital	American Broadcasting
95	Ashland Oil	Marathan Oil
96	Bank of America	Bank of California
97	Bendix	Ben Gay
98	Boise Cascade	Ben Gay
99	Canadian Pacific	Georgia Pacific
100	Ben Gay	Bendex
101	Champion	Chose Manhatten
102	Chrysler	John Deer
103	Coca Cola	Ala Moana
104	Clorex	Memorex
105	Chase Manhattan	Champion
106	Data General	General Steel
107	Cities Services	Citicorp
108	Canoca	Coca Cola
109	Southern Pacific	Georgia Pacific
110	Boeing	Boise Cascade
111	Ford	Fotomat

112	El Paso Gas	San Diego Gas
113	Union Oil of California	Bank of California
114	TRW	NCR
115	Citicorp	Cities Services
116	Revlon	Avon
117	Woolworth	Whirlpool
118	Rockwell	Honeywell
119	United Technologies	Threshold Technology
120	Fotomat	Ford
121	San Diego Gas	El Paso Gas
122	Southwestern Gas	San Diego Gas
123	Bell & Howell	Rockwell
124	Honeywell	Rockwell
125	Avon	Revlon
126	Threshold Technology	United Technology
127	Allied Corp.	Alles Chalmers
128	ACF Industries	AMF
129	Beckman Instruments	Becton, Dickinson
130	Big Three	Square D
131	Boston Edison	Con Edison
132	Brooklyn Union Gas	San Diego Gas
133	Bulova	Burroughs
134	Campbell Soup	Honeywell
135	Carolina Power & Light	Wisconsin Power & Light
136	Carpenter Technology	United Technology
137	AMF	ACF Industries
138	Becton Dickinson	Beckman Instruments
139	Carolina Freight Carriers	Carolina Power & Light
140	Central Maine Power	Central & South West
141	Clark Oil & Refining	Clark Equipment
142	Coldwell Banker	Rockwell
143	Colonial Store	Colt Industries
144	Commonwealth Edison	Con Edison
145	Central & South West	Central Soya
146	Consolidated Foods	Consolidated Freightway
147	Continental Can	Continental Air Lines
148	Colt Industries	Colonial Stores
149	Clark Equipment	Clark Oil & Refining
150	Central Soya	Central & South West
151	Continental Oil	Continental Telephone
152	Copper Range	Cooper Laboratories
153	Consolidated Freightway	Consolidated Foods
154	Continental Air Lines	Continental Can
155	Corning	Crocker
156	Cummins Engine	Combustion Engineering
157	Cooper Laboratories	Cooper Range
158	Cyprus Mines	Bucyrus-Erie
159	Del Monte Corp	Delta Air Lines
160	Continental Telephone	Continental Oil
161	Detroit Edison	Commonwealth Edison
162	Diebold	Di Giorgio
163	Dresser Industries	Dreyfess
164	Eastman Kodak	Eastern Airlines
165	Crocker	Corning

166	Bucyrus-Erie	Cyprus Mines
167	Emuson Electric	Emery Air Freight
168	Empire Gas	Emery Air Freight
169	Di Georgia	Georgia Pacific
170	Dreyfus	Dresser Industries
171	Faberge	Fairchild
172	Federal Mogul	Federal Paper Board
173	Fieldcrest	First Chicago
174	Fairchild	Faberge
175	Emery Air Freight	Emerson Electric
176	Ford Motor	Foremost
177	General Dynamics	General Food
178	Gerber	Gillette
179	Grand Union	Estern Union
180	Hushey	Hewlett-Packard
181	Hilton	Holiday Inn
182	Host	Houston Natural Gas
183	Inland Steel	Kaiser Steel
184	Internation Harvester	International Paper
185	Johnson & Johnson	Jonathan Logan
186	Kennecaft Copper	Kimberly Clark
187	Lone Star Gas	Lone Star Industries
188	Lukens Steel	Kaiser Steel
189	First Chicago	Fieldcrest
190	General Food	General Dynamics
191	Holiday Inn	Hilton
192	Foremost	Ford Motor
193	Hewlett Packard	Hershey
194	Houston Natural Gas	Host
195	International Paper	International Harverster
196	Johnathan Logan	Johnson & Johnson
197	Magnavox	Lenox
198	Lone Star Industries	Lone Star Gas
199	Federal Paper Board	Federal Mogul
200	Kimberly Clark	Kennecott Copper

```
+LILIST 1,9000

10  DIM W$(200,2),F$(3),P(4)
20 D$ = "": REM  CONTAINS A CONTROL-D
30 F$(1) = "COMPUTER":F$(2) = "STOCK":F$(3) = "AIRLINE"
40 P(1) = .99:P(2) = .95:P(3) = .90:P(4) = .85
45 M$ = "        "
50  PRINT "WHAT WORD DATABASE DO YOU WANT TO USE?": PRINT "1. COMPUTER":
     "2. STOCK": PRINT "3. AIRLINE"
60  INPUT A
70  PRINT D$;"OPEN ";F$(A);",V001"
80  PRINT D$;"READ ";F$(A)
90  FOR I = 1 TO 200: FOR J = 1 TO 2: INPUT W$(I,J): NEXT J: NEXT I
100  PRINT D$;"CLOSE "
105  PRINT "INPUT A SEED": INPUT IS
110  PRINT "WHAT ALGORITHM DO YOU WANT TO USE?"
120  INPUT B: IF B < 0 THEN  STOP
130  GOSUB 40000
140  CALL JTABLE + 15
150 VOC$ = "CLASSIC.VOC"
160  GOSUB 40100
170  PRINT "FIRST SOME PRACTICE WORDS"
180 I = 1
190  PRINT I,W$(I,1): GOSUB 40400: FOR Z = 1 TO 150: NEXT Z: GOSUB 300
200  IF I < 11 THEN  GOTO 190
210 Q = 0:S = 0:N = 0:T = 0
220  PRINT "WE WILL NOW START THE EXPERIMENT"
230 I = 1
240  PRINT I,W$(I,1): GOSUB 40400: FOR Z = 1 TO 150: NEXT Z: GOSUB 300
250  IF I < 201 THEN  GOTO 240
260  PRINT "THANK YOU FOR PARTICIPATING IN THE EXPERIMENT"
270 T = (200 - Q - S - N) / .2
280  PRINT B;T;"-";Q;"-";S;"-";N
290  GOTO 110
300  IF  RND (IS) > P(B) THEN  GOTO 320
310  PRINT M$;W$(I,1): GOTO 400
320 R =  RND (IS)
330  IF R < .33 THEN  GOTO 390
340  IF R < .66 THEN  GOTO 360
350  PRINT M$;"?":Q = Q + 1: GOTO 400
360 U =  INT ( RND (IS) * 200) + 1
370  IF U = I THEN  GOTO 360
380  PRINT M$;W$(U,1):N = N + 1: GOTO 400
390  PRINT M$;W$(I,2):S = S + 1
400  FOR Z = 1 TO 150: NEXT Z:I = I + 1: PRINT : RETURN
```

Appendix D

Code # ______________

CATEGORY [ ] AIRLINE

[ ] STOCK BROKER Age ________

[ ] COMPUTER

TIME SPENT AT A COMPUTER TERMINAL A DAY ____________ hrs.

1. Have you used voice recognition equipment before
   ☐ YES
   ☐ NO

2. Have you seen voice recognition equipment used before
   ☐ YES
   ☐ NO

3. Considering that you know how many typing errors you normally make, and considering the number of errors you saw the voice system make, how well do you like the voice system.

Really Dislike	Don't Like	Neutral: If I have it, fine If I don't have it, fine	Like It	Really Like It
☐	☐	☐	☐	☐

4. Is the accuracy of the voice recognition system adequate enough to make you want to use it in your daily job?

Very Inadequate	Slightly Inadequate	Neutral	Slightly Adequate	Very Adequate
☐	☐	☐	☐	☐

5. Comparing voice input to manual typing input, is voice input:

Undoubtedly Worse	Moderately Worse	The Same	Moderately Better	Undoubtedly Better
☐	☐	☐	☐	☐

6. Do you consider yourself a slow, intermediate, or fast typist?

Slow ☐ Intermediate ☐ Fast ☐

7. How accurately do you think you type?

Poor ☐ Fair ☐ Quite Well ☐

8. Did you ever take a typing course?

Yes ☐ No ☐

9. What part of the country, (USA) did you grow up in? ______________

10. How many years experience have you had in typing information into computers?

______________years.

11. Please tell the experimenter how much and what type of educational background you have.

APPENDIX E Raw Data

C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	42	22	20	10	12	35	1	1	1	4	5	4	13	2	2	1	1	1
	38	27	11	14	13	27	2	2	2	4	4	4	12	2	2	1	2	1
	36	17	19	12	5	32	1	2	2	3	2	2	7	2	3	2	1	14
	32	21	11	10	11	27	3	2	2	4	2	5	11	1	2	1	2	11
	32	21	11	15	6	28	3	2	1	4	4	4	12	1	2	1	3	1
	31	20	11	9	11	38	6	2	2	5	5	5	15	2	2	1	1	1
	30	16	14	9	7	34	1	1	2	4	4	4	12	2	2	1	1	0
	29	22	7	9	13	28	2	2	1	4	5	5	14	3	3	1	1	11
	29	18	11	12	6	38	2	2	2	5	4	5	14	1	3	2	4	10
	29	14	15	7	7	33	1	2	1	5	5	5	15	1	1	2	3	5
	29	15	14	9	6	27	2	2	1	5	4	5	14	2	3	2	1	4
	29	18	11	12	6	31	5	2	1	2	1	2	5	3	3	1	2	8
	28	16	12	8	8	33	2	2	2	3	5	5	13	3	3	1	3	5
	28	15	13	5	10	32	1	1	1	4	4	4	12	2	2	1	3	0
	26	20	6	6	14	29	2	1	1	4	4	3	11	2	2	1	3	1
	24	15	9	9	6	27	2	2	2	3	2	2	7	1	3	2	3	9
	24	20	4	7	13	29	2	1	1	4	2	4	10	2	2	1	2	10
	23	15	8	6	9	31	2	2	2	3	4	3	10	1	2	2	2	0
	23	16	7	4	12	28	5	1	1	2	2	1	5	2	3	1	2	12
	22	15	7	9	6	33	1	1	1	4	5	5	14	1	1	2	2	0
	22	12	10	7	5	30	1	2	2	5	5	2	12	3	3	1	3	2
	21	13	8	9	4	32	6	2	2	2	3	2	7	3	3	1	3	6
	21	17	4	4	13	34	1	2	1	5	5	5	15	2	2	2	1	2
	20	13	7	3	10	36	1	2	2	3	2	2	15	1	3	2	1	1
	20	13	7	3	10	26	1	2	2	3	4	5	12	2	2	1	3	0
	20	13	7	6	7	38	3	2	1	2	2	4	8	2	3	1	2	14
	19	12	7	7	5	29	0	2	2	4	4	4	12	2	2	1	5	0
	18	12	6	5	7	26	1	2	2	5	5	5	15	2	2	1	2	3
	17	11	6	5	6	35	0	2	1	4	4	4	12	1	2	1	3	6
	17	12	5	3	9	35	3	1	1	4	5	5	14	3	3	2	3	10
	17	[illegible]	3	5	9	30	1	2	1	4	5	4	13	2	2	1	1	2
	17	[illegible]	6	5	6	27	0	2	1	5	5	5	15	2	2	2	3	1
	16	8	8	6	2	39	2	2	2	4	5	4	13	1	2	1	6	0
	15	12	3	8	4	29	3	2	2	4	5	4	13	3	3	1	2	3
	15	11	4	6	5	35	0	2	1	4	4	4	12	2	1	2	1	1
	15	12	3	6	6	31	2	2	1	5	5	5	15	2	2	1	3	0
	14	10	4	8	2	33	2	2	2	3	5	4	12	3	3	1	2	3
	13	10	3	6	4	35	1	1	1	4	5	4	13	2	2	2	1	2
	13	10	3	6	4	32	0	2	1	3	4	4	11	1	1	1	5	3
	13	10	3	6	4	37	2	2	1	5	5	5	15	2	2	1	5	9
	12	6	6	2	4	26	2	2	1	4	4	4	12	2	3	1	3	4
	12	9	3	5	4	34	3	1	1	3	3	4	10	2	2	1	6	2
	12	9	3	3	6	27	1	2	1	5	5	5	15	2	2	2	5	4
	11	8	3	3	5	31	1	2	1	4	3	3	10	1	2	1	5	1
	11	8	3	4	4	35	1	2	2	2	2	2	6	2	2	2	6	1
	11	7	4	2	5	28	0	2	1	3	3	2	8	1	3	2	1	1
	10	7	3	3	4	38	1	2	1	4	4	5	13	1	2	2	3	0
	10	7	3	3	4	35	1	2	2	4	5	5	14	2	1	1	3	4
	10	7	3	3	4	36	0	2	2	5	5	5	15	3	3	1	5	9
	9	7	2	4	3	26	1	2	1	4	2	4	10	1	2	2	3	1
	9	7	2	4	3	29	1	2	2	5	5	5	15	3	3	1	3	10
	8	7	1	5	2	32	2	2	1	5	5	5	15	2	2	1	5	4
	8	7	1	5	2	30	1	2	1	5	5	5	15	1	1	2	3	5
	7	3	4	1	2	35	2	2	2	4	5	5	14	2	2	1	4	1
	7	5	2	2	3	35	2	2	1	5	5	5	15	3	3	1	3	15
	6	5	1	1	4	34	4	2	1	4	5	5	14	2	2	2	3	8
	5	3	2	2	1	29	1	2	2	5	3	4	12	2	3	1	3	3
	4	2	2	1	1	28	2	2	1	3	4	4	11	2	3	2	2	2
	4	3	1	2	1	29	1	2	2	4	3	2	9	2	3	1	3	12
	3	2	1	1	1	27	2	2	2	5	5	5	15	2	2	1	5	0

3	2	1	1	1	29	2	2	1	5	5	5	15	2	2	2	1	12
3	3	0	1	2	34	1	1	2	4	4	3	11	2	2	1	3	1
3	3	0	2	1	49	1	2	1	4	3	4	11	1	2	2	4	15
3	3	0	2	1	27	1	2	1	4	4	4	12	2	2	2	1	2
2	2	0	1	1	32	2	2	2	4	4	4	12	2	2	1	3	2
2	1	1	0	1	32	2	2	2	4	3	4	11	2	2	1	6	6
1	0	1	0	0	33	2	2	1	4	5	5	14	2	1	1	3	0
1	0	1	0	0	29	0	1	1	5	5	5	15	2	3	1	1	2
0	0	0	0	0	27	0	2	1	5	5	5	15	2	2	1	1	4
0	0	0	0	0	28	2	2	1	5	5	5	15	2	2	1	5	1
0	0	0	0	0	45	6	2	1	5	5	5	15	1	2	1	5	7
0	0	0	0	0	35	1	2	1	4	5	4	13	2	2	2	1	6
0	0	0	0	0	33	2	2	2	3	3	4	10	1	1	1	1	3
0	0	0	0	0	27	1	2	1	5	5	5	15	2	2	1	3	2
0	0	0	0	0	34	1	2	1	4	4	5	13	1	2	1	2	2
0	0	0	0	0	29	1	1	1	3	3	3	9	3	3	1	3	10
0	0	0	0	0	25	2	2	1	5	5	5	15	2	3	1	1	6
0	0	0	0	0	34	0	1	1	5	5	5	15	1	1	1	2	3

17.0	15.5	19.0	17.0	17.0	26.5	2	2	1	5	5	5	67.5	2	2	2	1	73.0
17.0	20.5	7.5	17.0	24.5	55.0	1	1	2	4	4	3	21.0	2	2	1	3	19.0
17.0	20.5	7.5	24.0	17.0	78.0	1	2	1	4	3	4	21.0	1	2	2	4	77.5
17.0	20.5	7.5	24.0	17.0	10.5	1	2	1	4	4	4	31.5	2	2	2	1	32.0
13.5	15.5	7.5	17.0	17.0	42.0	2	2	2	4	4	4	31.5	2	2	1	3	32.0
13.5	13.0	19.0	7.0	17.0	42.0	2	2	2	4	3	4	21.0	2	2	1	6	56.0
11.5	6.5	19.0	7.0	6.5	48.5	2	2	1	4	5	5	52.0	2	1	1	3	6.0
11.5	6.5	19.0	7.0	6.5	26.5	0	1	1	5	5	5	67.5	2	3	1	1	32.0
5.5	6.5	7.5	7.0	6.5	10.5	0	2	1	5	5	5	67.5	2	2	1	1	47.5
5.5	6.5	7.5	7.0	6.5	18.5	2	2	1	5	5	5	67.5	2	2	1	5	19.0
5.5	6.5	7.5	7.0	6.5	77.0	6	2	1	5	5	5	67.5	1	2	1	5	59.0
5.5	6.5	7.5	7.0	6.5	63.5	1	2	1	4	5	4	43.0	2	2	2	1	56.0
5.5	6.5	7.5	7.0	6.5	48.5	2	2	2	3	3	4	14.5	1	1	1	1	41.0
5.5	6.5	7.5	7.0	6.5	10.5	1	2	1	5	5	5	67.5	2	2	1	3	32.0
5.5	6.5	7.5	7.0	6.5	55.0	1	2	1	4	4	5	43.0	1	2	1	2	32.0
5.5	6.5	7.5	7.0	6.5	26.5	1	1	1	3	3	3	10.5	3	3	1	3	67.0
5.5	6.5	7.5	7.0	6.5	1.0	2	2	1	5	5	5	67.5	2	3	1	1	56.0
5.5	6.5	7.5	7.0	6.5	55.0	0	1	1	5	5	5	67.5	1	1	1	2	41.0

DISTRIBUTION LIST

	No. of Copies
Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	4
Dean of Research Code 012A Naval Postgraduate School Monterey, CA 93940	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	1
Professor G. K. Poock Code 55Pk Naval Postgraduate School Monterey, CA 93940	125